

Weekly Report

Yuxin Ma

10.01.2012 - 10.07.2012

Intro

From this week on the VisWeek 2013 project has officially started. I mainly did an initial survey on my current research direction.

Readings & Research

The following two ideas might be interesting points on applications of active learning-based visual analysis approaches:

1. Step-by-Step Model/Data Refinement in Data Analysis

- (a) **Motivation** The computation cost of similarity can sometimes be too high to get the accurate value, while they can be refined increasingly with an initial value with limited precision. Sometimes the similarity can only be a probability distribution like uncertainty data. In these situations, we can build an application to improve our result increasingly.
- (b) **Idea** We present a system that handles the similarity and model refinement of a particular data mining task. Take clustering as an example, in the initial clustering process, less accurate similarities will be used for the algorithm. Based on the clustering result, some of the similarities will be refined as the input of next clustering process. Here the process will be treated as an "growing tree" of the continuous adjustment process. We can decide which part of similarities will be further refined. At the same time, our approach can be applied on any kind of input data but not only similarities.
- (c) **Metaphor** This application can be considered as "planting trees in a data farm". Each refinement process is a "tree" in a farm which represents a specific refinement direction.
- (d) **Application Scenario** In my opinion, the best scenario should be uncertainty active learning algorithms(like "active uncertainty clustering or so"). The uncertain data can be refined in each iteration with both manual adjustment or queries from algorithms.
- (e) **Key Points**
 - i. Some further survey on specific classification and clustering algorithms is needed here. For a specific task(classification or clustering), if no present uncertainty active learning methods available, we might develop new ones that applies our application.

- ii. The visual design of "data farm" should be clear enough to support comparison between different processes as well as pruning and adjusting the growing process.

2. Visual Analysis of Model Mixture

- (a) **Motivation** In many data mining tasks, fusing results from multiple models is a widely-used technique to easily improve the final result. Adaboost is a common algorithm for constructing a strong-classifier by combining results of several weak classifiers. In the research field of recommender systems, fusion is regarded as a ultimate method for prediction precision. However, how to combine results from multiple models is still a problem.
- (b) **Idea** Visual analysis approach might be a way of solving this problem. The output of each model can be visually compared, and visualization can be used to show strong and weak points of each model. This application might be task-dependent, unlike Point 1. above.
- (c) **Application Scenario** Recommender system can be a place to apply this kind of result fusion. Each kind of prediction model has better prediction spaces, so visual analysis approach can be applied here to adjust parameters of fusion model.

For visual analysis of heterogenous data, I found a Douban dataset which contains user behavior data about movies, music and books.

Practice & Skills

None.

Miscellaneous

Building Visualization Platform During this week I spent two days trying to build some examples on openFrameworks and CUV library. I decide to build an visualization framework of my own with a common visualization framework(openFrameworks) and some high-performance data-processing frameworks including GraphLabAPI(for graph and multidimensional data) and CUV Library(a matrix computation library on GPGPU with some implementations of data mining and machine learning algorithms). It will be a long-term project as a leisure-time coding practice and fundamental framework for future works.

Plan for Next Week

- Survey on active and uncertainty learning algorithms;
- summarize the two ideas.